TRMM
Real-time Data Processing Documentation

Prepared by

Erich Franz Stocker
TRMM Project Data Manager

September 22, 1998
Version 1.0

Introduction

The Tropical Rainfall Measuring Mission (TRMM) real-time system is a standalone data processing and distribution component running with the TRMM Data and Information System (TSDIS). It has a reduced set of products. These products also contain fewer parameters. Finally, the real-time system does not service the same group of TSDIS science users. It seeks to service weather organizations and modelers.

Regular TRMM data products start out as 24 hour level 0 data sets advertised by the Sensor Data Processing Facility (SDPF) to TSDIS via the automated DAN mechanism. The TRMM real-time data is provided by SDPF as quick look products after each TDRS contact. This data is pushed by SDPF onto the real-time machine at TSDIS after each pass.

The objective for SDPF is to provide level 0 data to TSDIS for real-time processing within 1 hour of the completion of the satellite down-link 90% of the time. This is a best effort objective. TSDIS' objective is to produce the real-time products within 30 min of receiving the level 0 data from SDPF. Again this is a best effort objective not a fixed requirement. This ensures that 90% of the time, the oldest bit of the data is no more than 3hrs from the actual collection time.

Users of the TRMM real-time system are provided access to the username and password for the ftp account from which the real-time data products may be retrieved. Users must have received approval for use from Dr. Ghassem Asrar at NASA Headquarters.

Data Products

The data products for the real-time system are different than the standard TRMM products. First, all the products because of the quicklook nature of the level 0 data were produced using the predictive rather than the definitive ephemeris file. In addition, the geolocation information in the files is stored as a 2 byte scaled integer rather than as a float and so some precision is lost. Given that the predictive ephemeris was used and the resolution of the instrument, this change in geolocation precision is not a major issue.

The real-time products also have fewer parameters. Real-time parameters concentrate on providing information about the surface rain and reduce the amount of vertical information they provide. This means that each product is much smaller and more focused on the area of interest to the target community. Their size also makes distribution via ftp only possible. For a detailed explanation of the format and contents of the real-time products, readers are directed to the document entitled *TRMM Realtime System File Specification,* dated September 1, 1998. This document may be obtained in Adobe PDF format from the TRMM real-time homepage.

Six real-time data products are produced. All of the products are produced in HDF format. For the Visible and Infrared Spectrometer (VIRS) the 1B01 product is produced. This product is really the same as the standard TRMM product except is produced as a quicklook product for each TDRS contact and it uses the predicative rather than the definitive ephemeris file. A reduced level 1B11 product is produced for the TRMM Microwave Imager (TMI). Also for TMI, a reduced level 2A12 rain product is produced. Both TMI products have a suffix of *rt* to indicate that these are real-time and not standard TRMM products.

Three real-time products are produced for NASDA's Precipitation Radar (PR). These are parameter subsets of the standard TRMM products. Only level 2 PR products are part of the real-time set. The real-time 2A23 product provides rain-type information. Two real-time products are produced from 2A25. The 2A25r1 product contains near-surface rain information. The 2A25r2 product contains 20 levels of rain information. Each of these products is an HDF file and has the suffix *rt* to denote that it is the real-time version.

The naming convention used for TRMM real-time data products is different than those for standard TRMM products. The real-time system uses the naming convention:

Algorithm_id.SDPF_Level0_Date_Time_Production.rt

The algorithm_ids are 1B01 (VIRS level 1), 1B11 (TMI Level 1), 2A12 (TMI level 2 rain product), 2A23 (PR level 2 rain type product), 2A25r1 (PR near surface rain product), 2A25r2 (PR 20 level rain product). The SDPF_Level0_Date_Time_Production is the date and time when SDPF produced that product. All levels of real-time products use the date and time of their Level 0 product as provided by SDPF.

Standard TSDIS products contain the orbit number and collection date. These are not calculated as part of the real-time system.

All real-time products are compressed using the gzip utility and maintained for 3 days on-line. During this period they are accessible by authorized users. After 3 days they are archived to the TSDIS mass store device.

Level 0 Data Ingest and Pacor II

The TRMM real-time system gets its Level 0 and housekeeping data in a different manner from the TSDIS production system. As mentioned previously, TSDIS received a 24 hour Level 0 data for its regular products. The real-time system on the other hand receives a Level 0 product per instrument, instrument housekeeping and spacecraft housekeeping for each TDRS contact. This provides data of approximately one orbit or a little longer. Also, these products are not advertised to the real-time system via the automated Data Availability Notice (DAN) mechanism for the SDPF Data Distribution Facility (DDF) but instead are placed on to the TSDIS real-time machine directly by Pacor II. The ~16 contact Level 0 data files are equivalent to quicklook data in the regular TSDIS production.

L0 data provided by Pacor II maintain the SDPF naming convention. This includes the SDPF APID followed by the ISO date/time of processing. APID identifies the nature of the data. The TRMM real-time system handles the SDPF APIDs in Table one.

Table One
APID Processed by TRMM Real-time System

APID	INSTRUMENT	DATA TYPE
G045	ACS	ACS
G046	VIRS	VIRSHK
G047	TMI	TMIHK
G048	PR	PRHK
G051	VIRS	VIRS
G052	TMI	TMI
G053	PR	PR
G196	PR	IPSDUHA
G197	PR	IPSDUHB
G198	PR	IPSDUMA
G199	PR	IPSDUMB
G200	PR	IPSDULA
G201	PR	IPSDULB

Ingest Processing

Real-time system processing begins when Pacor II has completed transferring all required files. This is flagged when a DANxxxxx.done file appears in the directory to which Pacor II has transferred all the files. This DAN contains among other things the names of all the files that have been transferred to TSDIS.

An ingest daemon named getL0rt is constantly running on the real-time platform. Every 5 minutes it checks the SDPF directory to determine whether a DANxxxxx.done file has been received. If it has, it moves all the data to the level 0 directory (named in the environment variable RTLEVEL) and activates the L1A processing script for all three instruments. Each of these scripts in turn will activate the appropriate following processing step.

It also logs the name of the DAN and the arrival time, the data pool in which L1 and L2 products will be produced and activation time for the 3 L1A processing scripts. The log file name used for all logging by getL0rt is IngestLog.

Data Pools

Currently the TRMM real-time system has 12 data pools of 9 GB each. The ingest daemon decides which of these to use by looking for the one with the most available space. All real-time pools begin at the root – TRMMRT. The actual pool directories are named 01-12. The maximum number of pools available is maintained in the environment variable: RTDIRMAX.

Several times during the day, an archiving script is run that moves all the non TRMM real-time products to the mass store device. In addition, this script archives any TRMM real-time data products older than 3 days to the mass store device and removes all links to these files.

TRMM Real-time Product Processing

As indicated in the section on data ingest all processing is started by the ingest daemon – getL0rt. A graphical representation of the TRMM real-time processing flow is provided in figure 1.

Solid lines between rectangles in figure 1 indicate control flow. All rectangles are processes either C/FORTRAN programs or Perl scripts. Direction of arrow on the lines indicates caller and called processes. The rounded figure indicates data products. Dashed lines indicate real-time product output. Regular lines indicate data product access. Bold lines indicate the activation of script files. Arrows indicate whether a process reads data or writes data. An arrow from a product to a process indicates that the process is reading the data product. An arrow from a process to a product indicates the process is writing the product. Shadowed products indicate TRMM real-time products. Non shadowed products indicate data necessary for production of real-time products.

The level 1A scripts process in parallel. On wolkenbruch (the real-time machine) this means that each L1A process has its own processor. When the Perl scripts that activate the level 1A algorithms for VIRS, TMI and PR have successfully created the 1A data products (these are non-HDF products), they activate the Level 1B Perl scripts run1B01, run1B11, run1BPR respectively.

When the L1B01 (VIRS) code has successfully created the 1B01 HDF product, it is compressed using the gzip routine. A link is then placed in the ftp directory in the appropriate subdirectory. This gzipped 1B01 product is then available for user retrieval.

In parallel with run1B01, the Perl script run1B11 is run. This script provides three major services. It creates the full 1B11 (TMI) HDF product. When this has completed, it activates the Perl script that starts the 2A12 process. In parallel, run1B11 runs the algorithm that creates the extracted 1B11 real-time product. When this real-time parameter extraction has completed the 1B11 real-time product is compressed via gzip and a link to it is placed in the appropriate subdirectory of the ftp directory.

In parallel with the previously described 1B Perl scripts, run1BPR creates the 1B21 (PR) product. Upon successful creation of this HDF product, it activates the run2A21 Perl script and in parallel activates the run1CPR Perl script.

The run1CPR Perl script runs in parallel with the script generating the 2A21 product. When a 1C21 product has been successfully produced, the script activates the run2A23 Perl script.

The run2A23 Perl script executes the algorithm that generates the 2A23 HDF product. Upon successful completion it activates the Perl script that creates the 2A23 real-time product. In parallel it activates the Perl script that creates the 2A25 product.

The Perl script run2A23rt executes the algorithm that extracts the real-time parameters from the full 2A23 HDF product created in the previous step. Upon the successful creation of the 2A23 real-time product, it gzips the HDF real-time product. It then creates a link to the product in the appropriate subdirectory of the ftp directory.

The Perl script 2A25 performs a number of functions. It ensures that both its inputs 2A23 and 2A21 that are run in parallel have been successfully completed. It checks the process stream to ensure that no 2A23 or 2A21 products are running. If they are, the script goes to sleep for a pre-determined time and then checks again. It continues to do this until the process table is clear of 2A21 and 2A23. It then activates the 2A25 algorithm that creates the entire standard TRMM 2A25 with all 80 vertical levels. When the full 2A25 HDF product has been successfully completed, it activates in parallel, the two scripts that create the 2A25 run-time products.

The Perl script run2A25r1 executes the algorithm code that extracts the near surface rain information from the full product and creates the first 2A25 real-time product. It then gzips the real-time product and establishes a link to the compressed file in the appropriate ftp subdirectory. In parallel, the Perl script run2A25r2 executes the algorithm code that extracts 20 vertical levels from the 80 levels of the full 2A25 product. When it has successfully created the 20 level 2A25 real-time product. It gzips this second 2A25 real-time product and creates the appropriate link into the ftp directory.

Data Distribution

Only approved users have access to the TRMM real-time data. Access to the data requires a user name and password. The data distribution account is an ftp only account. Users can not telnet to the account. Passwords are changed at random intervals and approved users are informed of new passwords.

TRMM real-time data is distributed only via ftp. The working assumption is that users will routinely and probably in an automated fashion check the TRMM real-time ftp directory and retrieve new data. The real-time system provides no interface. It also does not notify users when data products become available. The entire distribution premise is that users will wish to automate the ftp process and execute their automated scripts at various times of the day.

Figure 1
TRMM Real-time Processing Flow

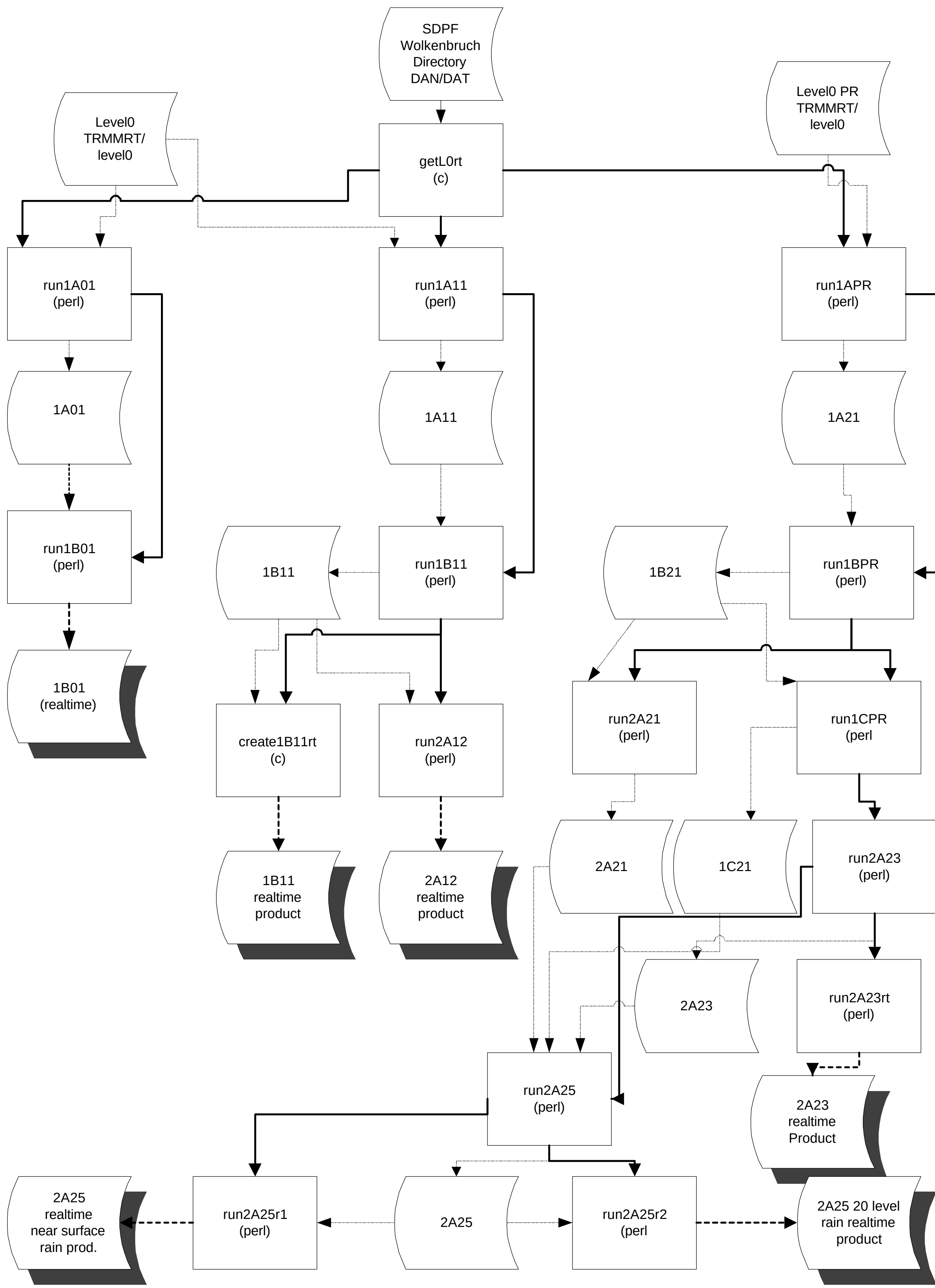
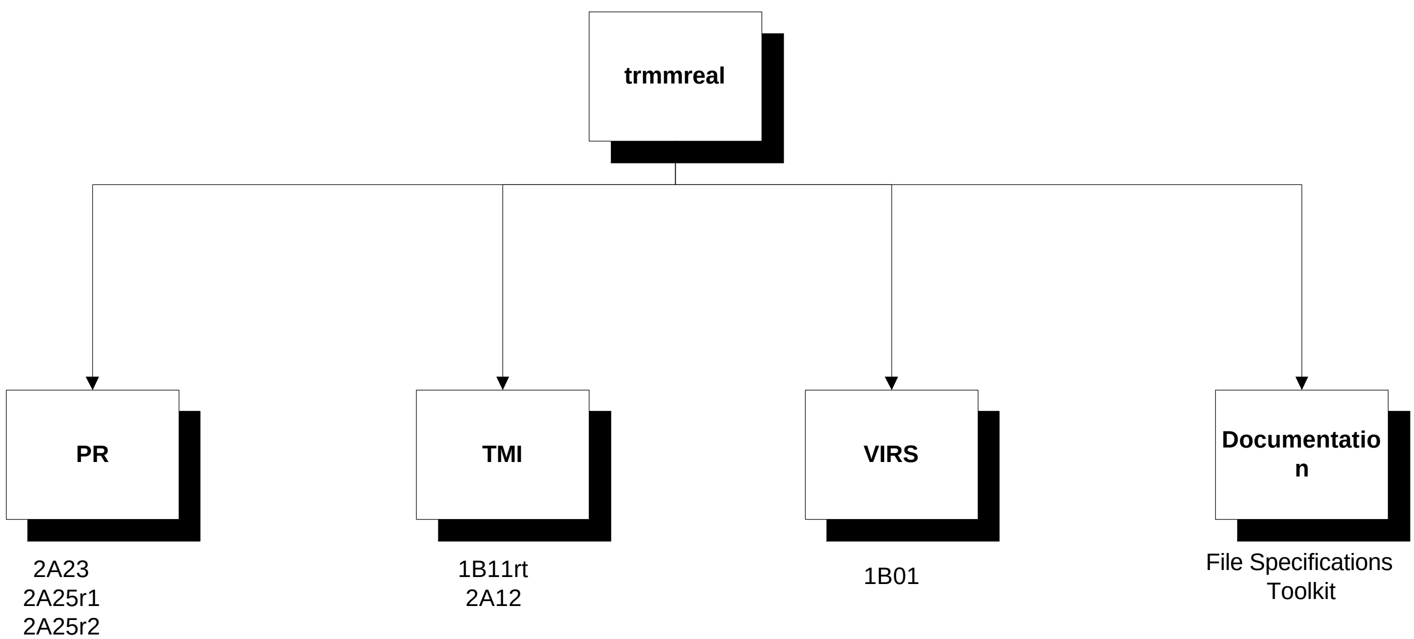


Figure 2 presents the directory structure of the TRMM real-time ftp account directory. A successful ftp connection puts the user in the top directory of the tree.

Figure 2
TRMM Real-time FTP Directory



Each subdirectory contains links to the actual location of the data products in the real-time data pools. Data products are maintained on-line for 3 days and then archived on the TSDIS mass storage device. As a result of processing lag, and sometimes maintenance needs, products may on occasion remain for longer than 3 days. However, this extended availability is not standard. Given the purpose of real-time data, it is expected that users will retrieve data as quickly as possible and as often as possible. After 48 hours from the date of collection the full TRMM products would be available to the science community through the GSFC DAAC. As a result, delayed retrieval of real-time products from the mass-store is not expected and no software exists to do it automatically.

Data Compatibility

TRMM real-time products are compatible with standard TRMM products. Standard TSDIS processing algorithms are used to create the data products. Two major exceptions exist. Instead of starting with a 24 hr. L0 data product, real-time processing starts with a QL product. Geolocation is not as accurate as a predicative vs. definitive empheris is used and gelocation parameters in the real-time products are 2 byte scaled integers rather than 4 byte float values. All real-time products, except 1B01 (VIRS) contain only a subset of the parameters included in the full TRMM standard product.

Standard TRMM algorithms provided by TRMM algorithm developers are used for producing 1B01, 1B11, 1B21, 1C21, 2A21, 2A23 and 2A25. 1B21, 1C21, and 2A21 are not real-time products but are required for generating 2A23 and 2A25. The full products are then read by extraction software that retrieves the desired parameters from the full product and copies them into the real-time product.

The TMI level 2A12 algorithm code is not the same as that used for generating the standard TRMM product. Because of the nature and processing demands of the standard algorithm, it was not possible to use it and still maintain the time objectives. As a result, the responsible scientist provided a separate real-time algorithm that generated the desired parameters within the established time objectives. He has agreed to maintain the integrity of this algorithm and ensure that it is compatible with the standard product as required.

With the exception of 2A12 whenever new versions of the science algorithms are received from the investigators, they are installed into the real-time system as well as into the TSDIS production system. This approach ensures that the algorithm quality control applied to the standard products is also included in the real-time products. This also allows the developer, with the exception of 2A12, to maintain a single version of an algorithm. Improvements and corrections to these algorithms are then automatically included in the real-time system.